Darshan M

Senior AI Engineer — Agentic Systems & LLM Applications
Bengaluru, India • +91 96205 77943 • darshan31597@gmail.com

SUMMARY

Senior software engineer with 7 years at Ivanti, the last 3+ focused on building production LLM and agentic AI systems. Designed and shipped a multi-tenant agentic platform on LangGraph and the Model Context Protocol (MCP), serving 10–15 enterprise customers with a fleet of specialized agents over a stateful, checkpointed graph runtime. First-place winner at DevCon 2024 (and a prior DevCon top-3 finalist) and an organization-wide MVP, with a consistent record of taking generative-AI ideas from POC to production at scale.

TECHNICAL SKILLS

Agentic AI & LLMs: Multi-agent orchestration, agent handoffs & separation-of-duties safety, LangGraph, LangChain, Model Context Protocol (FastMCP), RAG, LLM-as-judge evaluation & reranking, structured outputs, intent detection, NER, semantic search, text-to-SQL, human-in-the-loop, stateful checkpointing, prompt engineering (Chain-of-Thought, Self-Consistency)

LLM Platforms & Tooling: Azure OpenAI, Azure AI Foundry, Azure AI Agents (Bing grounding), OpenAI, GPT-4 Turbo, GitHub Copilot, text embeddings, OpenTelemetry tracing, token & cost monitoring

Testing & Evaluation: LLM-driven test automation, offline LLM evaluation, performance & load testing, regression testing, assertion-based validation

Languages: Python, C#, .NET / ASP.NET, C / C++, SQL

Backend & Data: FastAPI, microservices, REST / OData APIs, multi-tenancy, Redis, asyncio & concurrency, Pydantic v2, httpx, Tenacity, event-driven streaming (Redis pub/sub, Azure SignalR)

Cloud & DevOps: Microsoft Azure, Docker, Azure DevOps CI/CD, Kubernetes (Lens), New Relic APM, SAST/SCA (Polaris, Black Duck), Atlassian Confluence (Rovo MCP), UV monorepo

EXPERIENCE

Ivanti — Bengaluru, India

Aug 2019 – Present

Senior Software Engineer · AI & Agentic Systems *Feb 2023 – Present*

Intelligent Agent — production multi-tenant agentic AI platform for IT Service Management. Python 3.13 · LangGraph · FastMCP / MCP · Azure OpenAI · FastAPI · Redis · Docker.

- Architected a multi-agent system on LangGraph with 9+ specialized subgraphs (intent detection, incident creation, service-request automation, knowledge RAG, conversational Q&A, summarization) coordinated by a central intent-router with conditional routing and human-in-the-loop interrupts — in production for **10–15 enterprise customers**.
- Designed and built both sides of a production MCP deployment: a FastMCP server exposing 10+ tools (NL-to-SQL, semantic and full-text search, service-catalog and asset lookup, grounded web search) and 15+ reusable prompts, plus a connection-pooled, async-safe MCP client gateway with HTTP/stdio transports and per-tenant header injection.
- Built stateful, resumable conversations using LangGraph checkpointing backed by Redis (TTL session expiry, stale-checkpoint sweep), enabling multi-turn flows that survive restarts and scale horizontally across stateless pods.
- Implemented a knowledge-retrieval (RAG) pipeline: query decomposition → parallel multi-tool search (OData + full-text) → LLM-as-judge semantic reranking → grounded answer generation with KB-resolution validation and escalation routing.
- Engineered end-to-end multi-tenancy and strict data isolation — SID parsing, session validation, Redis-backed session cache with fail-open fallback, and tenant-scoped state, configuration, and logging — enforcing an AuthContext on every tool, prompt, and resource call.
- Streamed real-time agent status, reasoning steps, and final responses to Microsoft Teams over a Redis pub/sub event bus; deployed custom models on Azure AI Foundry with OpenTelemetry tracing and token-usage monitoring; and shipped via multi-stage Docker builds, New Relic APM, and Azure DevOps CI/CD gated by SAST/SCA scans.

Agent Test Harness — async, multi-user, LLM-driven performance and end-to-end correctness suite for the Intelligent Agent. Python 3.13 · asyncio · httpx · redis.asyncio · Azure OpenAI.

- Built an async, multi-user harness that provisions synthetic users and drives realistic conversations with an LLM, gating every streamed Redis event with fail-fast assertions across 4 agent types (incident, service request, knowledge, Q&A) — replacing unrepeatable manual Teams testing and cutting QA functional-regression time.
- Reported per-node and per-agent latency (avg / min / max and end-to-end wall time) to catch execution-time regressions when developers change graph nodes.
- Extending the harness with an LLM-as-judge offline evaluator to flag response-quality regressions, and a per-node latency benchmark that requires developer sign-off when a node runs slower than its baseline.

DevAgent — multi-agent AI developer associate that carries an Azure DevOps work item from triage through fix, standards review, security gate, and test validation. Python · MCP · GitHub Copilot · Azure DevOps · Atlassian Rovo · Black Duck · Polaris.

- Architected a 4-agent pipeline (Vanguard, Sentinel, Black Duck Agent, Tester) that takes an Azure DevOps work item from triage → root-cause fix → standards review → security gate → quality judgment, with the developer supplying only a PAT token and committing the final change.
- Vanguard ranks assigned work items by a composite urgency heuristic (priority, severity, target release, escalation signals), detects repo ownership by content scan, and applies minimal root-cause fixes; Sentinel independently reviews each diff against company standards pulled from Atlassian Confluence via the Rovo MCP, looping findings back by severity until the change is review-clean.
- Black Duck Agent gates the change on Black Duck (SCA) and Polaris (SAST) scans; the Tester agent then runs concurrent plain-English regression cases against a live tenant as an LLM-as-judge, returning per-step pass/fail verdicts.
- Enforced separation of duties as a safety model — no agent reviews its own work or pushes code — with explicit, auditable handoffs and human-in-the-loop approval, all built on MCP to stay model-agnostic and portable across repositories.
- Won **1st place and a $20,000 award** at DevCon 2024 for “CoPilot,” turning plain-language requirements into ITSM business objects via Chain-of-Thought and Self-Consistency prompting (**months of configuration to minutes**); an earlier top-3 DevCon idea became the POC that established Ivanti’s permanent AI ITSM team.
- Built a real-time natural-language analytics interface (Azure SignalR + GPT-4 Turbo) that turns plain-English questions into interactive dashboards — applying NER to identify and match business objects and fields, Levenshtein-distance entity mining, and on-the-fly SQL construction — sharply improving responsiveness for analytical queries.
- Upgraded entity recognition with precomputed Ada text embeddings and semantic-similarity matching, aligning semantically related terms beyond literal string matches for more accurate, context-aware results.

Software Engineer *Aug 2021 – Feb 2023*

- Drove technical convergence after Ivanti’s acquisition of Cherwell: built connectors to transform and merge metadata-driven data between CSM and ISM, and led a cross-geo (India–UK) team as interim lead of the India “Cyborg” squad.

Associate Software Engineer *Aug 2019 – Jul 2021*

- Built RESTful OData APIs using the Microsoft EDM library to dynamically map database tables to entities, expanding the product’s query and integration surface.
- Supported a multi-tenant push-notification system (GCM, Azure Notification Hub, Amazon SNS) and automated role-based object permissions by mining business-object usage across workspaces, reports, and forms.

AWARDS & RECOGNITION

- DevCon 2024 — 1st Place & $20,000 award for “CoPilot” (AI-driven ITSM configuration from natural language).
- Most Valuable Professional (MVP), organization-wide across all Ivanti product teams — recognized with a 10,000-share equity grant and one-on-one reviews with the CEO and CTO.
- DevCon — Top-3 idea for natural-language-to-SQL in ITSM.
- Best Performer / Individual Contributor (3×) and Best Team (2×).

EDUCATION

Bachelor of Engineering — University Visvesvaraya College of Engineering (UVCE), Bengaluru *2015 – 2019*